Supplementary materials

Article type: Brief Communication

Title: Low cell number ChIP-seq reveals chromatin state-based regulation of gene transcription in the rice male meiocytes

Authors: Aicen Zhang[1,#], Hanli You [2,4,#], Lei Cao[2,4,#], Yining Shi[1,#], Jiawei Chen[2,4], Yi Shen[2], Shentong Tao[1], Zhukuan Cheng[3*], Wenli Zhang[1*]

[1]State Key Laboratory for Crop Genetics and Germplasm Enhancement, CIC-MCP, Nanjing Agricultural University, No.1 Weigang, Nanjing, Jiangsu 210095, P. R. China. [2]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. [3]Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops, Yangzhou University, 225009 Yangzhou, China. [4] University of Chinese Academy of Sciences, Beijing 100049, China.

Materials and methods

Collection of rice male meiocytes at the early prophase I stage

Germinated seeds of Zhongxian 3037 (*indica*) were grown in the rice paddy field in Changping District, Beijing, China (116.42°, 40.10°) as described before (Jiang et al., 2020). The germ cells at the early meiotic prophase I stage were manually

collected using capillary collection method as previously described (Chen and Retzel, 2013; Jiang et al., 2020). Briefly, the anthers with meiocytes at the early prophase I were collected and dissected out from approximately 3–4 mm long spikelets using dissecting needles. Worm-like meiotic cell clusters were gently squeezed out from the anther wall and released into one drop of 1× phosphate-buffered saline (PBS). Pure meiocytes were manually collected into 1.5 ml Eppendorf tube using capillary glass pipettes under microscope. Approximately 10,000 cells were cross-linked using 1% of formaldehyde (v/v) in 1× PBS at 4°C for 10min. The excessive formaldehyde was quenched by using a final concentration of 0.125M glycine for an additional 5min. The cross-linked cells were washed twice with 1× PBS and store at -80°C for later use.

Low cell number ChIP-seq (LCNChIP-seq) and data analyses

10,000 cross-linked PMC pellet was suspended in pre-cooled cell lysis buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 0.6% NP-40) for lysis on ice for 10min. After centrifuge at 12,000 rpm at 4℃ for 5min, the cell pellet was resuspended in the cell lysis buffer and incubated for additional 5 min on ice. After washing twice using 1xMNase digestion buffer (20 mM Tris-HCl, pH 7.5, 15 mM NaCl, 60 mM KCl, 1 mM $CaCl_2$), the lysates were trimmed using MNase (Cat# M0247S, New England Biolabs) for 10 min at 37°C. After adding 2x stopping buffer (100 mM Tris-HCl, pH 8.0, 20 mM EDTA, 200 mM NaCl, 2% Triton X-100, 0.2% sodium deoxycholate), MNase trimmed chromatin was sonicated for 10 cycles (20s on, 40s off). After

centrifuge twice at 12,000 rpm at 4℃ for 5min each, the clean supernatant was incubated with anti-H3K27me3 (Cat# 9733, Cell Signaling Technology) or anti-H3K4me3 (Cat# ab8580, abcam) with constant rotation at 4℃ overnight. The remaining procedures for recovery of ChIPed DNA were conducted following the published procedures (Saleh et al., 2008). Each antibody was biologically replicated. Biologically replicated ChIP-seq and input libraries were prepared using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (NEB, E7645S). All prepared libraries were sequenced on the Illumina NovaSeq platform.

Raw reads of all sequencing data were trimmed using fastp (Chen et al., 2018). Reads with quality value Q < 25 and read length < 50 bp were discarded. The clean reads were mapped to the rice MSU7.0 reference genome (http://rice.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_7.0/all.dir/) with Bowtie2 aligner (Langmead and Salzberg, 2012). Picard MarkDuplicates (http://broadinstitute.github.io/picard/) was used to remove any PCR duplicates. Samtools (Li et al., 2009) was used to select reads with mapping quality (MapQ) higher than 30 for further analyses. Deeptools (Ramirez et al., 2016) was used to convert the bam files to the bigwig files with the bam files normalized by RPKM (Reads Per Kilobase per Million mapped reads). Read distributions across the whole genome were visualized by using IGV (Thorvaldsdottir et al., 2013). The fastq files of H3K4me3 data from maize meiocytes at the early prophase I stage were download from Bioproject PRJNA185817 (Kianian et al., 2018) and mapped to *Zea mays* AGPv4 reference genome following the steps

mentioned above.

Association of H3K4me3 and H3K27me3 peaks with genes and phasiRNA encoding loci

To reduce impacts of sequencing depth on peak calling, we down-sampled the ChIP-seq reads from each biological replicate of H3K4me3 or H3K27me3 to the same amount by using samtools (Li et al., 2009). MACS2 software (Zhang et al., 2008) was used for H3K4me3 and H3K27me3 peak calling with parameter as below: '--broad' parameter on (-f BAM -g 3.73134e+08 --nomodel -q 0.01 --broad --broad-cutoff 0.1) for H3K27me3 peak calling and off (-f BAM -g 3.73134e+08 --nomodel -q 0.01) for H3K4me3 peak calling. Only peaks present in both replicates were kept for downstream analyses. Bedtools (Quinlan and Hall, 2010) was used to associate H3K4me3 and H3K27me3 peaks with genomic loci and genes (from the 2 kb upstream of the TSS to 2 kb downstream of the TTS). H3K4me3 peaks with consecutive peak length no less than 800 bp (representing the top 10% length of all peaks) were regarded as broad H3K4me3 peak. Any genes overlapping H3K4me3 or H3K27me3 peaks were defined as peak-related genes. The phasiRNA encoding loci were obtained from the public data (Jiang et al., 2020; Li et al., 2020). A custom script was used to calculate normalized read counts of related genes and genomic loci.

Construction of regulatory network

The input FPKM matrix of gene expression from the early prophase I of rice

meiocytes and other rice tissues was downloaded from the published literature (Jiang et al., 2020) and Rice Expression Database (RED) (Xia et al., 2017). For construction of co-expression related regulatory network, we first used the R package WGCNA (v1.69) (Langfelder and Horvath, 2008) for weighted gene co-expression network analyses. Genes with top 15,000 median absolute deviation (MAD) value were used for construction of the co-expression network. According to the ME (module eigengene) values, which represent the overall gene expression pattern in each co-expressed gene module, and the number of TFs within each individual module, we selected the turquoise module containing genes preferentially expressed at the early prophase I of rice meiocytes for further construction of key TF centered gene regulatory network. The Perl package TF_collaborative (https://github.com/hwei0805/TF_CollaborativeNet) was used for the inference of regulatory network using the matrix of gene expression in the turquoise module. The top 100 high-confidence target genes for each TF were predicted, and the overlaps between TF targeted genes were calculated, all TFs were finally clustered according to the triple-link algorithm (Nie et al., 2011). TF centered collaborative subnetworks were finally generated to represent different biological functions using the aforementioned Perl package.

Comparation of the expression levels of genes in meiocytes at the early prophase I stage in *Arabidopsis*, rice and maize

The RNA-seq data from meiocytes at the early prophase I stage in *Arabidopsis* and

maize were obtained from the previous studies (Dukowic-Schulze et al., 2014; Walker et al., 2018). The clean reads were mapped to TAIR10 and *Zea mays* AGPv4 reference genome using Hisat2 (Kim et al., 2015), respectively. StringTie (Pertea et al., 2015) was used to calculate the expression level (RPKM, Reads Per Kilobase per Million mapped reads) of each gene. Genes with RPKM >= 1 were considered as expressed ones. The homologous gene pairs between rice and *Arabidopsis*/maize were identified using reciprocal BLAST hit between rice and *Arabidopsis*/maize proteins as previous reported (Zhang et al., 2015). All genes in each species were divided into four subgroups according to their expression values (high, middle, low and no expression). Three subtypes of genes were correlated with H3K4me3/H3K27me3, including Group 1(G1) with genes highly expressed in all species examined; Group 2 (G2) with genes highly expressed in only one species; Group 3 (G3) with genes lowly or non- expressed in one species but highly expressed in the other two species.

GO enrichment analyses

Gene Ontology (GO) enrichment analyses were conducted in Plant Regulomics (Ran et al., 2020). A *p*-value less than 0.05 was used to screen significantly enriched GO terms. Redundant GO terms were filtered out by using the REVIGO website (Supek et al., 2011).

Discussion

The traditional chromatin immunoprecipitation and sequencing (ChIP-seq) is a powerful and robust methodology for global profiling of chromatin modifications across a wide range of eukaryotic species (Du et al., 2013; Kimura, 2013; Liu et al., 2014; Liu et al., 2016; Wang et al., 2021; Zhang et al., 2021). But, it usually starts with 1-10M or even more cells, which are much more than the number of cells used in micro-scale ChIP, like a micro-scale ChIP and sequencing method (μChIP-seq) (Dahl et al., 2016), STAR ChIP-seq (small-scale TELP-assisted rapid ChIP-seq) (Zhang et al., 2016), and LCNChIP-seq (used in our study). As compared to the traditional ChIP-seq, any robust micro-scale ChIP, including LCNChIP-seq, can be a good choice for profiling of histone modifications in plants or tissues with very limited amount available for use like embryo in rice and *Arabidopsis*, epical meristem cells, or specific cell types with very limited numbers like male or female meiocytes at each specific meiotic stage. This will save time and labors for collection of materials for ChIP. Moreover, it is worth noting that the quality of antibody will be one of the key factors affecting the success of micro-scale ChIP. In general, the higher the quality of the antibody used, the higher the success of the micro-scale ChIP experiment. In addition, more effort needs to be invested to add more epigenomic marks, such as DNA methylation and H3K9me2, two important heterochromatic marks during meiosis (Choi et al., 2018; Wang et al., 2022, He et al., 2022) using low cell number or even at a single cell level, which will advance our understanding of epigenomic regulation in the plant meiosis. ScWGBS (sing-cell Whole Genome Bisulfite Sequencing, using 50 cells) and bisulfite-converted randomly integrated fragments

sequencing (BRIF-seq) have been successfully applied to profile DNA methylation genome-wide in rice gametes and zygote (Zhou et al., 2021) and the microspore of maize (Li et al., 2019), respectively. Thus, it is necessary and promising to adapt either of methods for profiling of DNA methylation in the rice meiocytes at prophase I stage using low cell numbers or a single cell.

References

Chen, C. and Retzel, E.F. (2013) Analyzing the meiotic transcriptome using isolated meiocytes of *Arabidopsis thaliana*. *Methods Mol. Biol.* **990**, 203-213.

Chen, S., Zhou, Y., Chen, Y. and Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890.

Choi, K., Zhao, X., Tock, A.J., Lambing, C., Underwood, C.J., Hardcastle, T.J., Serra, H., Kim, J., Cho, H.S., Kim, J., Ziolkowski, P.A., Yelina, N.E., Hwang, I., Martienssen, R.A. and Henderson, I.R. (2018) Nucleosomes and DNA methylation shape meiotic DSB frequency in *Arabidopsis thaliana* transposons and gene regulatory regions. *Genome Res.* **28**, 532-546.

Dahl, J.A., Jung, I., Aanes, H., Greggains, G.D., Manaf, A., Lerdrup, M., Li, G., Kuan, S., Li, B., Lee, A.Y., Preissl, S., Jermstad, I., Haugen, M.H., Suganthan, R., Bjørås, M., Hansen, K., Dalen, K.T., Fedorcsak, P., Ren, B. and Klungland, A. (2016) Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **537**, 548-552.

Du, Z., Li, H., Wei, Q., Zhao, X., Wang, C., Zhu, Q., Yi, X., Xu, W., Liu, X.S., Jin, W.

and Su, Z. (2013) Genome-wide analysis of histone modifications: H3K4me2, H3K4me3, H3K9ac, and H3K27ac in *Oryza sativa* L. *Japonica*. *Mol. Plant* **6**, 1463-1472.

Dukowic-Schulze, S., Sundararajan, A., Mudge, J., Ramaraj, T., Farmer, A.D., Wang, M., Sun, Q., Pillardy, J., Kianian, S., Retzel, E.F., Pawlowski, W.P. and Chen, C. (2014) The transcriptome landscape of early maize meiosis. *BMC Plant Biol.* **14**, 118.

He, C., Chen, Z., Zhao, Y., Yu, Y., Wang, H., Wang, C., Copenhaver, G.P., Qi, J. and Wang, Y. (2022) Histone demethylase IBM1-mediated meiocyte gene expression ensures meiotic chromosome synapsis and recombination. *PLoS Genet.* **18**, e1010041.

Jiang, P., Lian, B., Liu, C., Fu, Z., Shen, Y., Cheng, Z. and Qi, Y. (2020) 21-nt phasiRNAs direct target mRNA cleavage in rice male germ cells. *BMC Plant Biol.* **11**, 5191.

Kianian, P.M.A, Wang, M., Simons, K., Ghavami, F., He, Y., Dukowic-Schulze, S., Sundararajan, A., Sun, Q., Pillardy, J., Mudge, J., Chen, C., Kianian, S.F., Pawlowski, W.P. High-resolution crossover mapping reveals similarities and differences of male and female recombination in maize. *Nat. Commun.* **9,** 2370.

Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. methods* **12**, 357-360.

Kimura, H. (2013) Histone modifications for human epigenome analysis. *J. Hum. Genet.* **58**, 439-445.

Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**, 559.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. methods* **9**, 357-359.

Li, C., Xu, H., Fu, F.F., Russell, S.D., Sundaresan, V. and Gent, J.I. (2020) Genome-wide redistribution of 24-nt siRNAs in rice gametes. *Genome Res* **30**, 173-184.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The Sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079.

Li, X., Chen, L., Zhang, Q., Sun, Y., Li, Q. and Yan, J. (2019) BRIF-seq: Bisulfite-converted randomly integrated fragments sequencing at the single-cell level. *Mol. Plant* **12**, 438-446.

Liu, N., Fromm, M. and Avramova, Z. (2014) H3K27me3 and H3K4me3 chromatin environment at super-induced dehydration stress memory genes of *Arabidopsis thaliana*. *Mol. Plant* **7**, 502-513.

Liu, X., Wang, C., Liu, W., Li, J., Li, C., Kou, X., Chen, J., Zhao, Y., Gao, H., Wang, H., Zhang, Y., Gao, Y. and Gao, S. (2016) Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* **537**, 558-562.

Nie, J., Stewart, R., Zhang, H., Thomson, J.A., Ruan, F., Cui, X. and Wei, H. (2011) TF-Cluster: a pipeline for identifying functionally coordinated transcription factors

via network decomposition of the shared coexpression connectivity matrix (SCCM). *BMC Syst. Biol.* **5**, 53.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T. and Salzberg, S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotech.* **33**, 290-295.

Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842.

Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F. and Manke, T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160-165.

Ran, X., Zhao, F., Wang, Y., Liu, J., Zhuang, Y., Ye, L., Qi, M., Cheng, J. and Zhang, Y. (2020) Plant Regulomics: a data-driven interface for retrieving upstream regulators from plant multi-omics data. *Plant J.* **101**, 237-248.

Saleh, A., Alvarez-Venegas, R. and Avramova, Z. (2008) An efficient chromatin immunoprecipitation (ChIP) protocol for studying histone modifications in *Arabidopsis* plants. *Nat. protoc.* **3**, 1018-1025.

Supek, F., Bosnjak, M., Skunca, N. and Smuc, T. (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one* **6**, e21800.

Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178-192.

Walker, J., Gao, H., Zhang, J., Aldridge, B., Vickers, M., Higgins, J.D. and Feng, X.

(2018) Sexual-lineage-specific DNA methylation regulates meiosis in *Arabidopsis.* *Nat. Genet.* **50**, 130-137.

Wang, L., Zheng, K., Zeng, L., Xu, D., Zhu, T., Yin, Y., Zhan, H., Wu, Y. and Yang, D.L. (2022) Reinforcement of CHH methylation through RNA-directed DNA methylation ensures sexual reproduction in rice. *Plant Physiol.* **188**, 1189-1209.

Wang, M., Li, Z., Zhang, Y., Zhang, Y., Xie, Y., Ye, L., Zhuang, Y., Lin, K., Zhao, F., Guo, J., Teng, W., Zhang, W., Tong, Y., Xue, Y. and Zhang, Y. (2021) An atlas of wheat epigenetic regulatory elements reveals subgenome divergence in the regulation of development and stress responses. *Plant Cell* **33**, 865-881.

Xia, L., Zou, D., Sang, J., Xu, X., Yin, H., Li, M., Wu, S., Hu, S., Hao, L. and Zhang, Z. (2017) Rice Expression Database (RED): An integrated RNA-Seq-derived gene expression database for rice. *J. Genet. Genomics* **44**, 235-241.

Zhang, A., Wei, Y., Shi, Y., Deng, X., Gao, J., Feng, Y., Zheng, D., Cheng, X., Li, Z., Wang, T., Wang, K., Liu, F., Peng, R. and Zhang, W. (2021) Profiling of H3K4me3 and H3K27me3 and their roles in gene subfunctionalization in allotetraploid cotton. *Front. Plant Sci.* **12**: 761059.

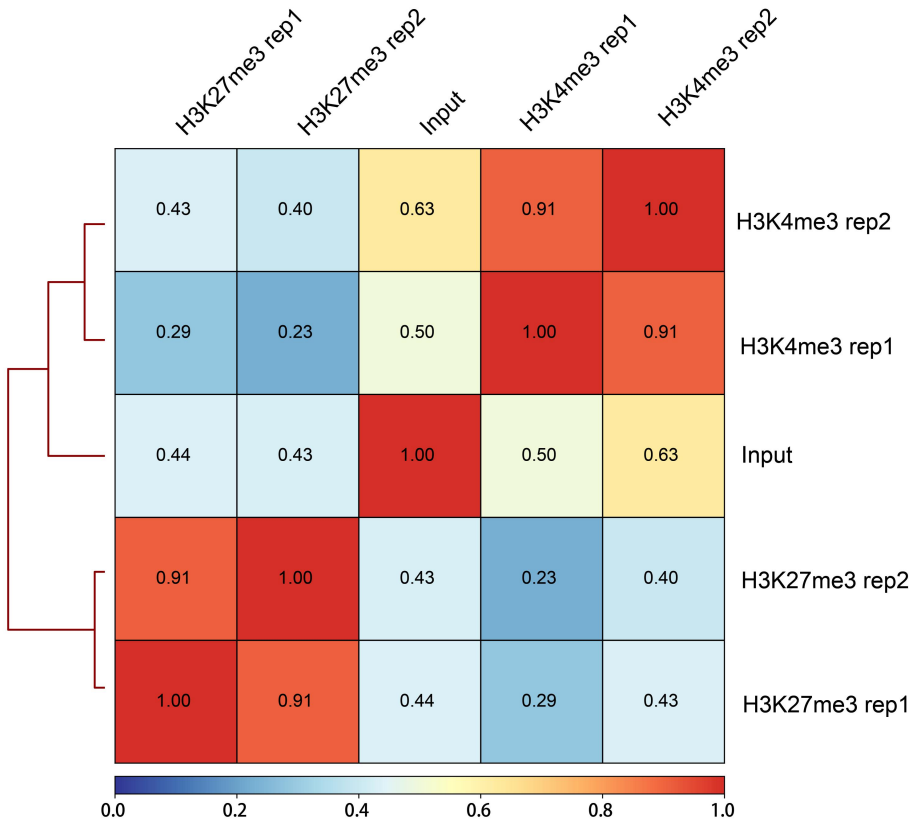
Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q., Liu, W., Kou, X., Zhao, Y., He, W., Li, C., Chen, B., Li, Y., Wang, Q., Ma, J., Yin, Q., Kee, K., Meng, A., Gao, S., Xu, F., Na, J. and Xie, W. (2016) Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553-557.

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C.A.,

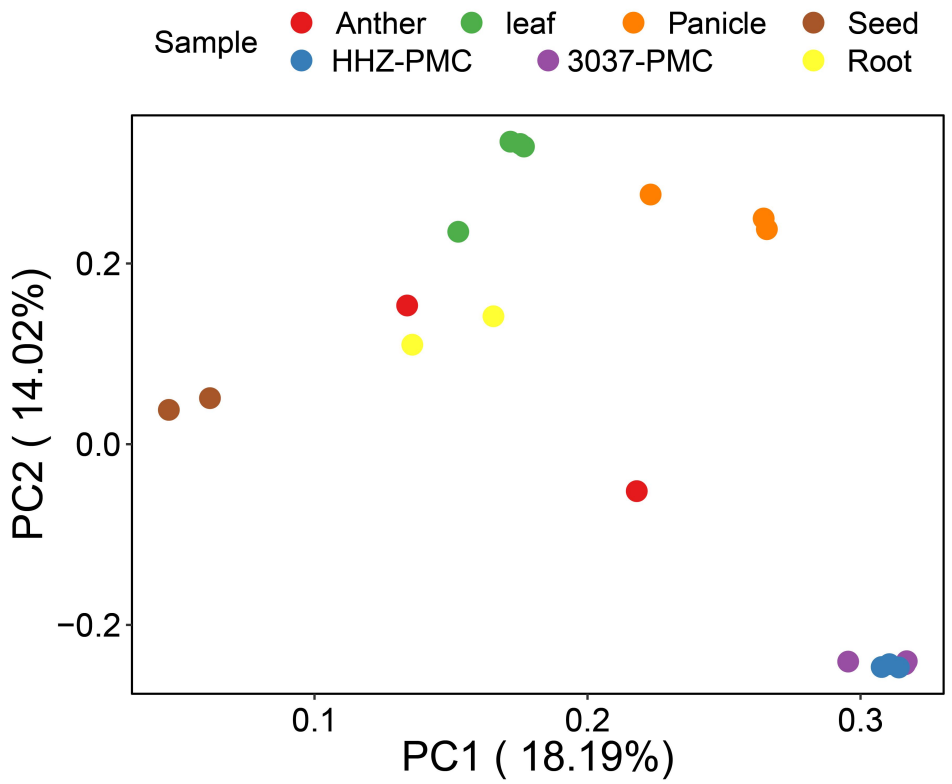
Scheffler, B.E., Stelly, D.M., Hulse-Kemp, A.M., Wan, Q., Liu, B., Liu, C., Wang, S., Pan, M., Wang, Y., Wang, D., Ye, W., Chang, L., Zhang, W., Song, Q., Kirkbride, R.C., Chen, X., Dennis, E., Llewellyn, D.J., Peterson, D.G., Thaxton, P., Jones, D.C., Wang, Q., Xu, X., Zhang, H., Wu, H., Zhou, L., Mei, G., Chen, S., Tian, Y., Xiang, D., Li, X., Ding, J., Zuo, Q., Tao, L., Liu, Y., Li, J., Lin, Y., Hui, Y., Cao, Z., Cai, C., Zhu, X., Jiang, Z., Zhou, B., Guo, W., Li, R. and Chen, Z.J. (2015) Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotech.* **33**, 531-537.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. and Liu, X.S. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137.

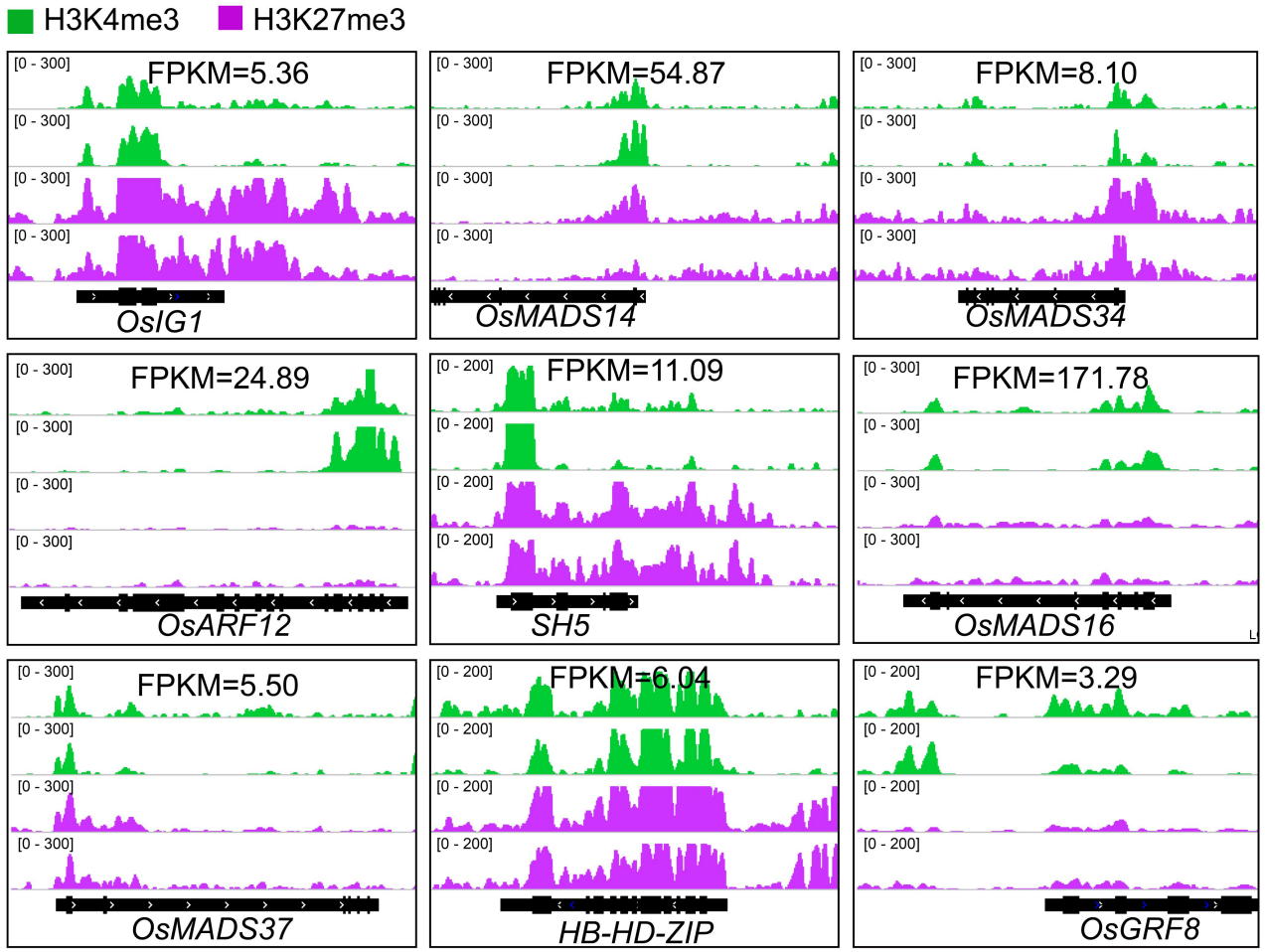
Zhou, S., Li, X., Liu, Q., Zhao, Y., Jiang, W., Wu, A. and Zhou, D.X. (2021) DNA demethylases remodel DNA methylation in rice gametes and zygote and are required for reproduction. *Mol. Plant* **14**, 1569-1583.



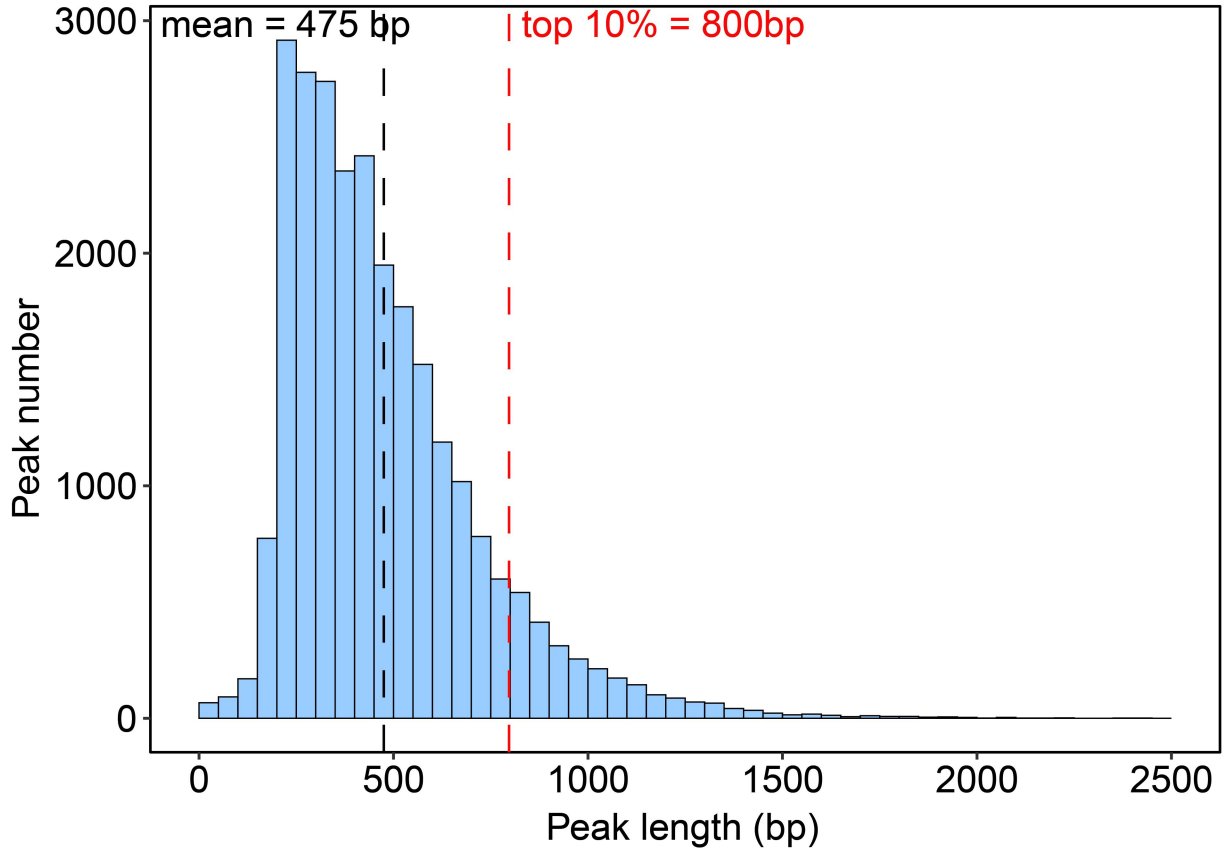
Supplementary Figure S1. Correlation analyses of biologically replicated H3K4me3 and H3K27me3 ChIP-seq datasets. Pearson correlation coefficient was shown.



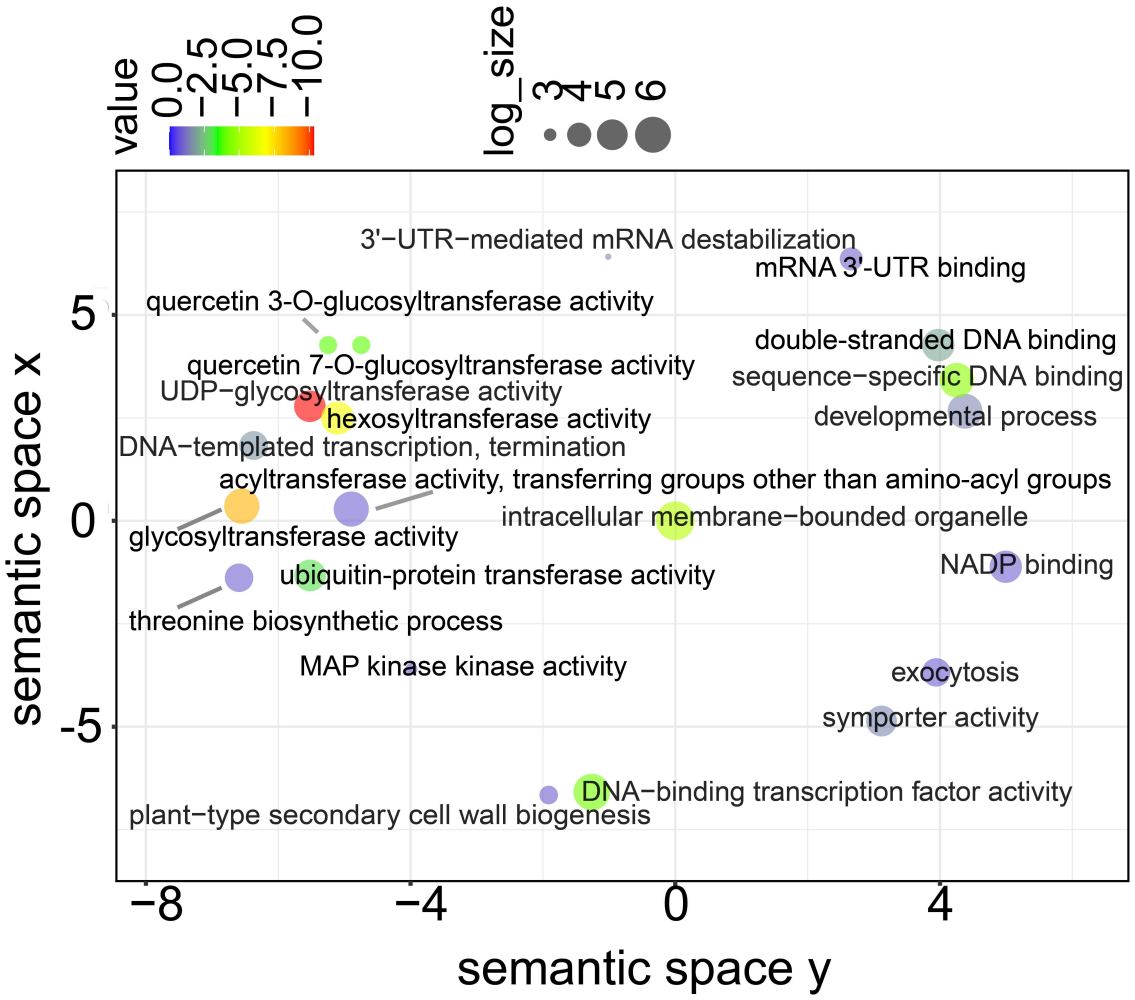
Supplementary Figure S2. Principal Component Analysis (PCA) of RNA-seq datasets generated from different rice tissues (HHZ represents the rice variety Huanghuazhan).



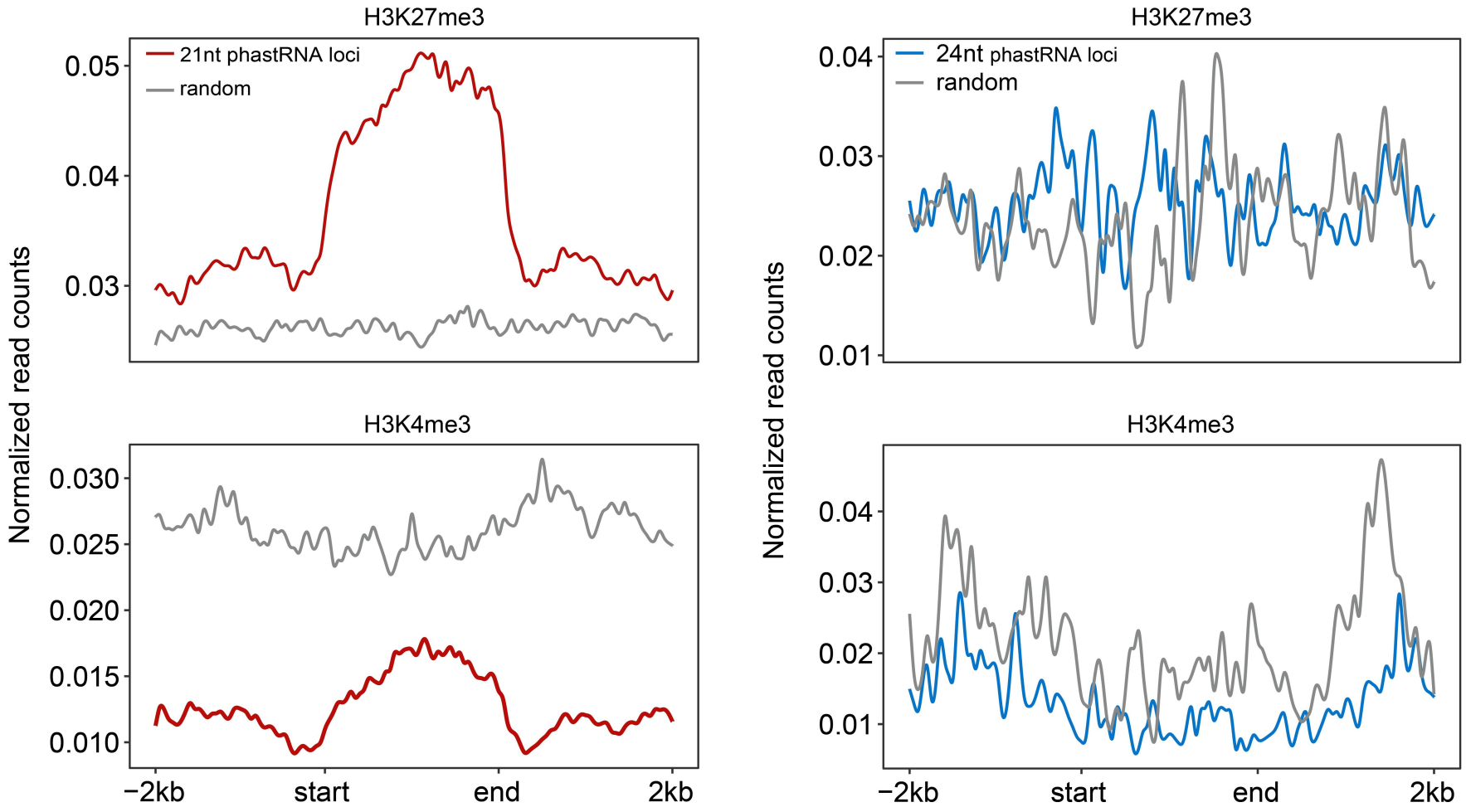
Supplementary Figure S3. IGV illustrating enrichment of H3K4me3 and H3K27me3 in 9 TFs in the network.



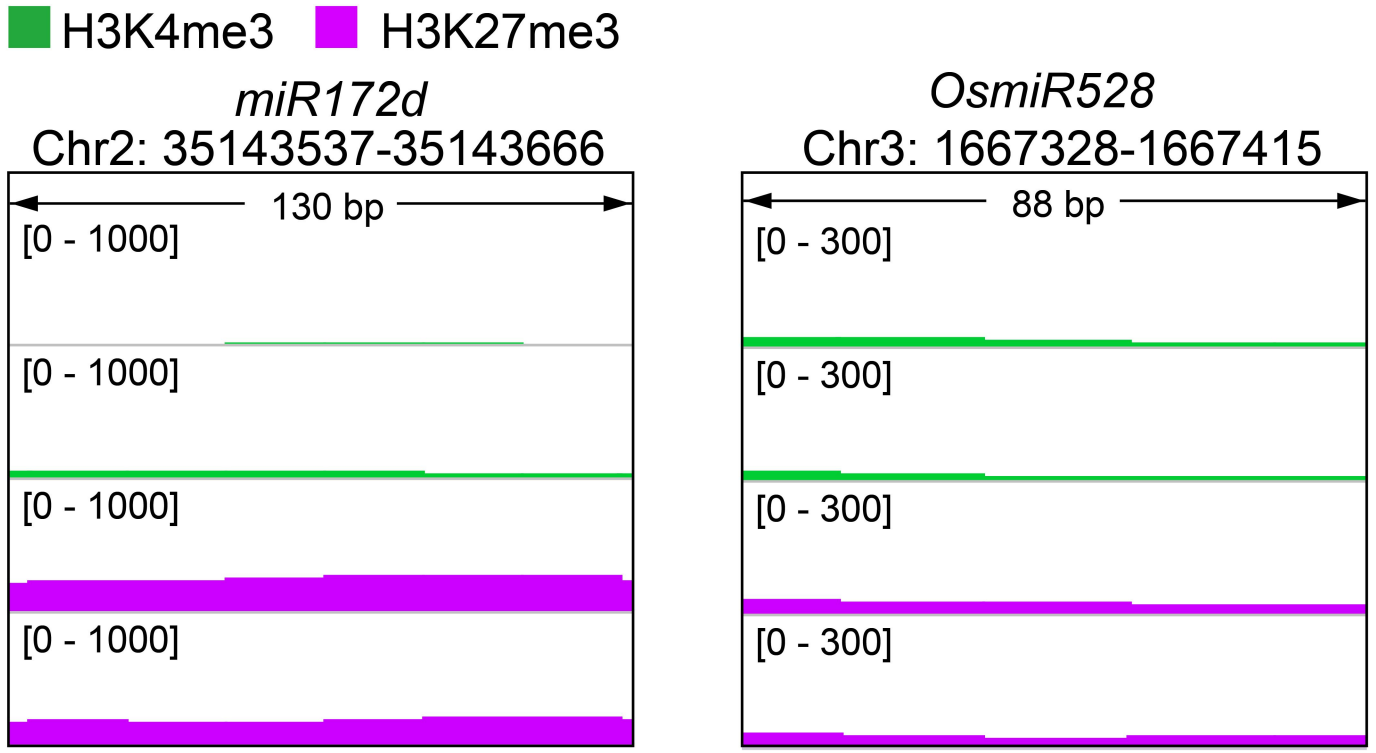
Supplementary Figure S4. Barlpot showing the length distributions of H3K4me3 peaks. The black dotted line reprsents the mean length of all peaks, the red dotted line represents the threshold of broad H3K4me3 peaks.



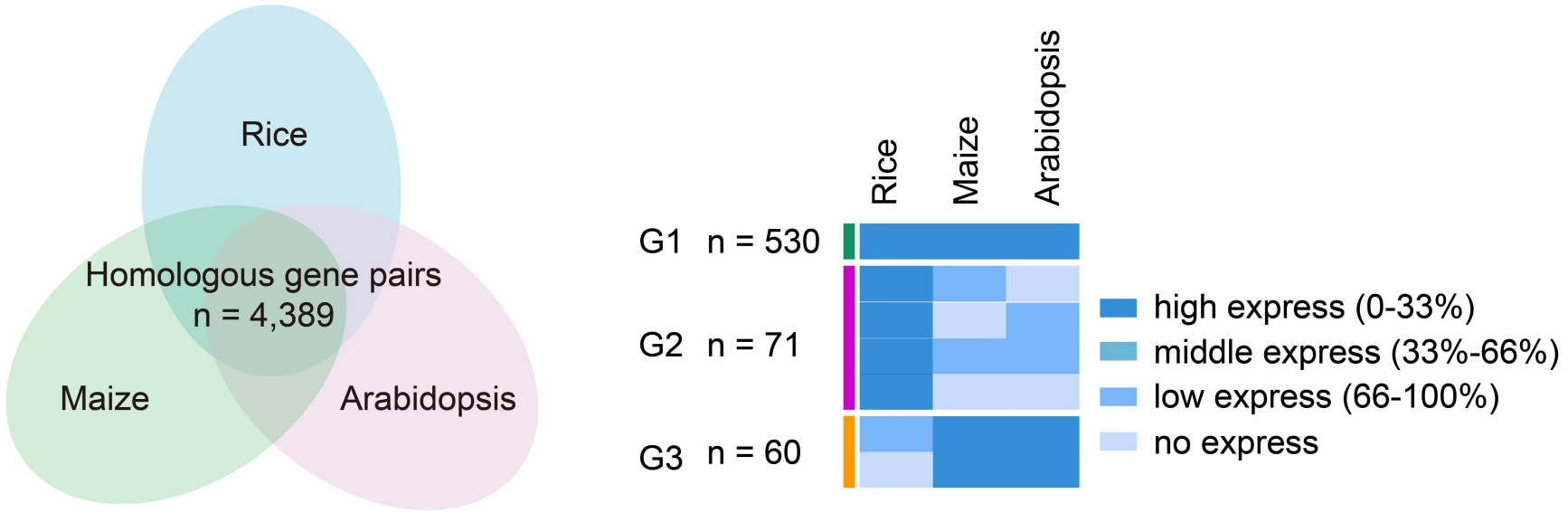
Supplementary Figure S5. GO term enrichment analyses of genes with broad H3K4me3.



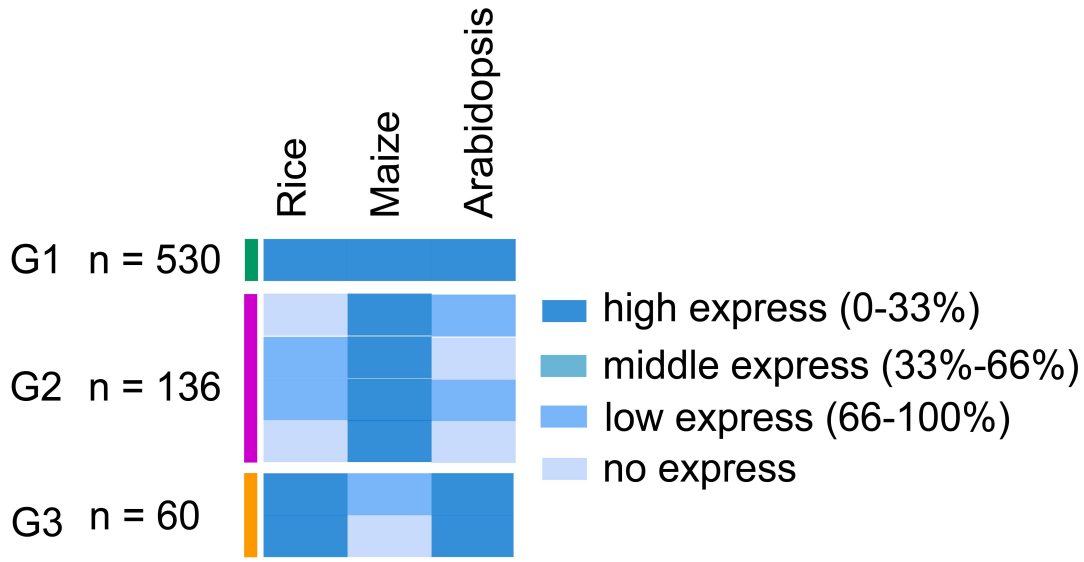
Supplementary Figure S6. Profiling of normalized H3K4me3 and H3K27me3 read counts from 2 kb upstream of the start point to 2 kb downstream of the end point of 21-nt phasiRNA encoding loci (left) and 24-nt phasiRNA encoding loci (right), respectively, which were reported in the rice gametes (Li et al., 2020).

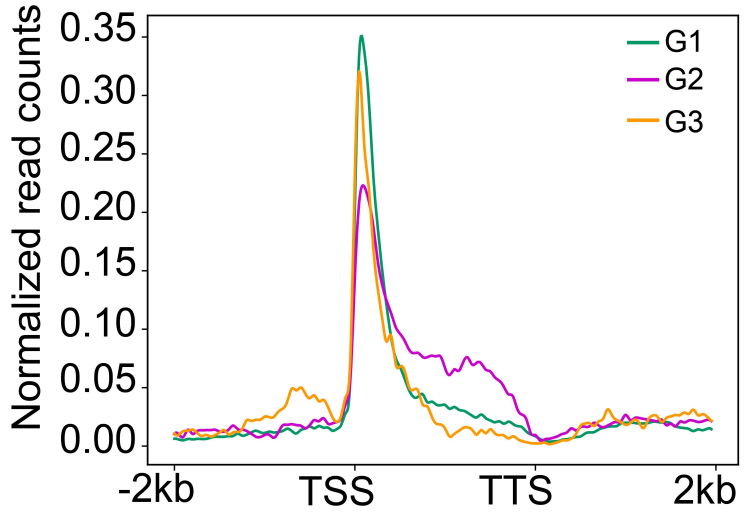


Supplementary Figure S7. IGV snapshots showing abundance of H3K4me3 and H3K27me3 in *miR172d* and *OsmiR528*.



Supplementary Figure S8. Homologous gene pairs in three plant species. Venn plot showing overlaps of homologous gene pairs with the identity of protein sequences greater than 50% among *Arabidopsis*, maize and rice. (**B**) Heatmap showing three distinct subtypes of homologous gene pairs, G1 (n = 530) with genes highly expressed in three species examined, G2 (n = 71) with genes highly expressed in rice only and G3 (n = 60) with genes highly expressed in *Arabidopsis* and maize.





Supplementary Figure S9. Homologous gene pairs in the maize meiocytes at early prophase I. Heatmap showing three distinct subtypes of homologous gene pairs, G1 (n = 530) with genes highly expressed in three species examined, G2 (n = 71) with genes highly expressed in rice only and G3 (n = 60) with genes highly expressed in *Arabidopsis* and maize. The enrichment levels of H3K4me3 across ± 2 kb from the TSSs to the TTSs of genes in G1, G2 and G3, respectively.